

This document is an English translation of the abstract of the KR144633 (INST ADVANCED ENGINEERING) APR. 21, 1998.

Abstract

For a method of compressing data using Lempel Ziv Welch (LZW), it is possible to increase compressibility for compressing 2 bytes Korean alphabet by re-organizing dictionary tree and inner table used for LZW compression algorithm for compressing 1 byte English alphabet.

(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) Int. Cl.⁶
G06F 17/22
G06F 17/28

(11) 등록번호 특0144633
(24) 등록일자 1998년04월21일

(21) 출원번호	특 1994-039802	(65) 공개번호	특 1996-025207
(22) 출원일자	1994년 12월 30일	(43) 공개일자	1996년 07월 20일
(73) 특허권자	고등기술연구원 연구조합 김준성		
(72) 발명자	서울시 중구 남대문로 5가 541 대우센터 빌딩 서두원		
(74) 대리인	서울시 중구 남대문로 5가 중앙사서함 2849호 박희진, 박영우		

심사관 : 이은철

(54) LZW를 이용한 완성형 한글의 데이터 압축방법

요약

LZW를 이용한 데이터 압축 방식에서, 일반적으로 사용되는 1바이트 영문전용 압축방식을 변형하여, 2바이트 한글에 대하여도 높은 데이터 압축률을 가질 수 있도록 LZW 압축 알고리즘에 사용되는 딕셔너리 트리(dictionary tree)와 내부 테이블을 재구성하여 압축률을 증가 시킨다.

대표도

도 1

명세서

[발명의 명칭]

LZW를 이용한 완성형 한글의 데이터 압축방법.

[도면의 간단한 설명]

제1도는 본 발명의 LZW를 이용한 데이터 압축방법의 흐름선도.

제2도의 a는 종래의 LZW를 이용한 데이터 압축방법의 기본 알고리즘의 초기 내부 테이블 구조 도해도.

제2도의 b는 본 발명의 LZW를 이용한 데이터 압축방법의 한글인식 알고리즘의 초기 내부 테이블 구조 도해도.

[발명의 상세한 설명]

본 발명은 LZW(Lempel Zive, Welch)를 이용한 데이터 압축 방법에 관한 것이며, 특히, LZW 압축 기법에 한글인식 과정을 추가시켜, LZW 압축 알고리즘에 딕셔너리 트리(dictionary tree), 즉 내부 테이블을 재구성시켜 압축률을 향상시킬 수 있는 데이터 압축 방법에 관한 것이다.

일반적으로, 데이터 전송속도를 향상시킬 수 있는 데이터 압축 방법에는 허프만 부호화 기법, 산술 부호화 기법과 LZW 압축기법이 있는데 그중 가장 많이 사용되고 있는 LZW 압축기법은 산술한 두가지 기법에 사용되는 입력 데이터의 확률 통계적인 성질에 의존하지 않고, 화일을 한번만 읽어서 처리하도록 하여 실시간 환경에서 좀더 효율적으로 적용하게 할 수 있는 유니버설 데이터 압축 기법이다.

상기 LZW 압축방법은 스트링 테이블(string table)이라는 번역 테이블을 만들어 이를 이용하는 방법으로, 논리적으로는 딕셔너리 트리(dictionary tree)라는 구조를 만들어, 각 기호열을 관리하게 되는데, 문자열을 이 트리의 한노드에 할당하고, 이 번지만을 전송하게 함으로써 압축이 이루어지는 방법이다.

그러나, 산술한 종래의 LZW 압축방법에서는 한글 인식 과정이 없이 모두 8비트 심볼 단위로 처리를 하였기 때문에, 2바이트로 구성되는 한글의 특성을 살리지 못하고, 압축과정을 수행함으로써, 압축률을 높일 수 있는 요소가 배제되는 문제점이 있었다.

본 발명은 산술한 문제점을 해결하기 위해, 종래의 LZW 압축방식에 완성형 한글인식 및 처리과정을 추가시킴으로써, 데이터 압축율을 높여서, 데이터 전송속도를 향상시키는 것을 목적으로 한다.

상기 목적을 달성하기 위해, 본 발명은 LZW 압축기법에 한글인식 과정을 추가시켜, 영문전용의 1바이트가 아닌 한글용 2바이트를 하나의 심볼로 인식하게 하고, LZW 압축 알고리즘에 사용되는 딕셔너리 트리과 내부 테이블을 재구성 시키는 방법을 특징으로 한다.

이하, 첨부된 도면으로 본 발명을 더욱 상세하게 설명하기로 한다.

제1도는 본 발명의 LZW 압축방법을 이용한 데이터 압축방법의 흐름선도로서, 그 첫 번째 단계는, 7비트 ASCII 코드용 128개의 노드와 완성형 한글 코드용 2350자를 딕셔너리 트리에 초기화 시키는 단계(S1-

2)로서, 이는 종래의 방법에서, 비트 부호의 모든 가능한 256개의 노드를 딕셔너리 트리에 초기화 시키는 단계(S1-1)를 대체한 것으로 된다. 그 다음 단계부터는 종래의 방법에서와 동일한 단계를 거치게 되는데, 즉, 입력문자를 읽어 스트링 w를 취하는 단계(S2)와, 다음 입력문자가 존재하는가를 판단하는 단계(S3)로 진행되는데, 만약 입력문자가 존재하지 않으면, 전단계(S2)의 스트링에 할당된 부호어를 출력하며(S10), 입력문자가 있으면, 다음 입력 문자를 읽어 캐릭터(k)로 설정하는 단계(S4)로 진행된다. 다음에 스트링과 캐릭터가 딕셔너리에 존재하는가를 판단하며(S5), 존재하면 스트링과 캐릭터를 새로운 스트링으로 하여(S6) 다시 상술한 S4단계로 진행된다. 만약에, 스트링과 캐릭터가 딕셔너리 트리에 존재하지 않으면, 스트링(w)에 할당한 부호어를 출력하며(S7), 계속하여, 딕셔너리 트리에 대한 새로운 노드를 생성하여 부호어를 할당시키며(S8), 다음에 캐릭터(k)를 새로운 스트링으로 취하며(S9), 다시 전단계(S4)로 귀환한다. 즉, 본 발명의 흐름선도에서는 점선으로 표시된 종래의 초기화 단계(S1-1)를 단계(S1-2)로 대체하는데 단계(S1-1)에서는 한글 존재의 여부를 판단하지 않고 모든 문자를 8비트 단위로 처리하였으나, 본 발명은 단계(S1-2)를 이용하여 더 높은 압축률을 가진다.

제2도의 (a)는 종래의 LZW를 이용한 데이터 압축 방법의 기본 알고리즘의 내부 테이블 구조의 도해도이며, 제2도의 (b)는 본 발명의 LZW를 이용한 데이터 압축방식의 한글 인식 알고리즘의 내부 테이블 구조 도해도로서, 테이블의 총크기를 8192로 하는 LZW 알고리즘에서의 내부 테이블 구조를 각각 나타낸다.

제2도의 (a)에서, 0에서 255까지의 부호어는 ASCII 코드로, 256 내지 511까지의 부호어는 9비트 부호화 영역으로, 512 내지 1023까지의 부호어는 10비트 부호화 영역으로, 1024 내지 2047까지의 부호어는 11비트 부호화 영역으로 하며, 2048 내지 4095까지의 부호어는 12비트 부호화 영역으로 하며, 4096 내지 8191까지의 부호어는 13비트 부호화 영역으로 한다. 반면에 제2도의 (b)에 따른 본 발명의 알고리즘의 내부 테이블 구조에서는, 딕셔너리 트리 테이블을 생성하는 단계에서, 제2도의 (a)의 초기에 256개의 ASCII 코드노드를 사용하는 대신 128개의 ASCII 코드와 2350자의 한글 코드용 노드를 딕셔너리 트리에 초기화시켜 사용한다. 즉, 제2도의 (a)에서의 9비트, 10비트, 11비트 부호화 영역 및 12비트 부호화 영역 중 일부를 본 발명에서는 한글 코드가 차지하게 된다. 즉, KSC 완성형 코드에 대한 한글 데이터 처리를 위하여 KSC 5601에 정의된 한글2350자를 초기 테이블 노드에 할당함으로써, 초기 테이블 노드는 ASCII 코드를 위한 128자+2350자인 2478자가 되도록 한다.

압축진행시 1617개의 부호어는 12비트로 하고, 나머지는 13비트로 출력하게 하여, 완성형 한글 인식 알고리즘의 내부 테이블 구조를 완성시킨다.

따라서, 변경된 초기 테이블 노드를 이용하여 LZW 알고리즘에 따라 압축을 행하게 한다.

이상, 본 발명에 따른 데이터 압축 방법에서는 영문 1바이트 대신에 2바이트 한 음절을 하나의 심볼로 간주하여 더 긴 비트 크기의 문자열에 대해 부호어를 할당시킴으로, 압축율의 향상을 가져온다.

(57) 청구의 범위

청구항 1

데이터 전송속도를 향상시키기 위한 데이터 압축용 LZW 압축 방법에서, 7비트 ASCII 코드용 노드와 한글 코드용 노드를 딕셔너리 초기에 초기화 시키는 단계와, 입력문자를 읽어 스트링(w)을 취하고, 다음 입력 문자를 읽어 캐릭터(k)를 취하는 단계와, 상기 스트링(w)과 캐릭터(k)가 딕셔너리 트리에 존재하는가를 판단하여, 존재하면, 스트링과 캐릭터를 새로운 스트링으로 하고, 존재하지 않으면, 스트링 w에 할당한 부호어를 출력하여, 딕셔너리 트리에 대한 새로운 노드를 생성하여 부호어를 할당한 후, 캐릭터(k)를 새로운 스트링으로 취하는 단계를 포함하는 LZW를 이용한 데이터 압축방법.

청구항 2

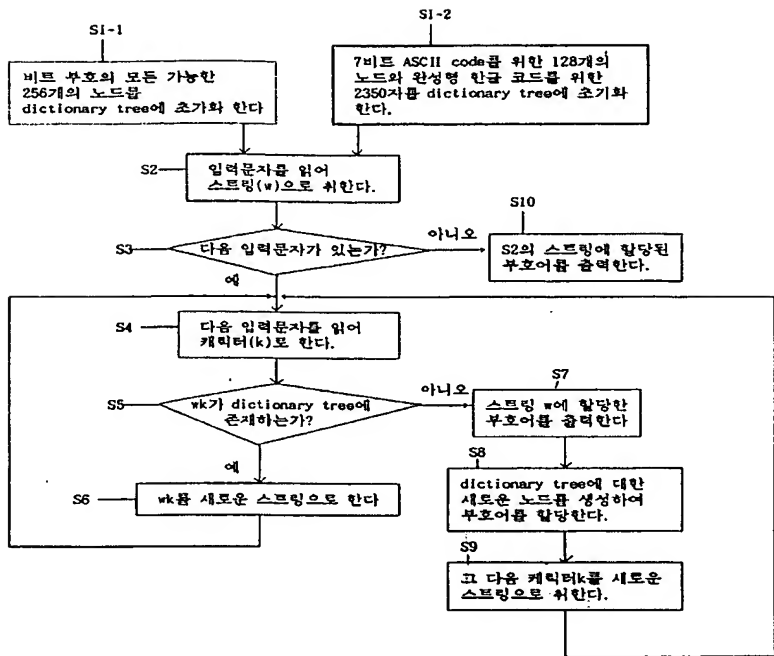
제1항에 있어서, 상기 7비트 ASCII 코드용 노드와 한글 코드용 노드를 딕셔너리 초기에 초기화 시키는 단계에서 7비트 ASCII 코드용 노드는 128개이며, 한글 코드용 노드는 2350개인 것을 특징으로 하는 LZW를 이용한 데이터 압축방법.

청구항 3

제1항 또는 제2항에 있어서, 상기 한글 코드용 노드를 위해 2바이트 한글 코드용 압축 방법을 이용하는 것을 특징으로 하는 LZW를 이용한 데이터 압축방법.

도면

도면1



도면2a

0~255	ASCII code
256~511	9비트 부호화 영역
512~1023	10비트 부호화 영역
1024~2047	11비트 부호화 영역
2048~4095	12비트 부호화 영역
4096~8191	13비트 부호화 영역

도면2b

0~127	ASCII code
128~2477	한글코드(2350자)
2478~4095	12비트 부호화 영역
4096~8191	13비트 부호화 영역